



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

# Identification of MAZ as a novel transcription factor regulating erythropoiesis

### Citation for published version:

Deen, D, Butter, F, Holland, ML, Samara, V, Sloane-Stanley, JA, Ayyub, H, Mann, M, Garrick, D & Vernimmen, D 2020 'Identification of MAZ as a novel transcription factor regulating erythropoiesis'. <https://doi.org/10.1101/2020.05.10.087254>, <https://doi.org/10.1101/2020.05.10.087254>

### Digital Object Identifier (DOI):

[10.1101/2020.05.10.087254](https://doi.org/10.1101/2020.05.10.087254)  
[10.1101/2020.05.10.087254](https://doi.org/10.1101/2020.05.10.087254)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Identification of MAZ as a novel transcription factor regulating erythropoiesis

**Darya Deen<sup>1</sup>, Falk Butter<sup>2</sup>, Michelle L. Holland<sup>3</sup>, Vasiliki Samara<sup>4</sup>, Jacqueline A. Sloane-Stanley<sup>4</sup>, Helena Ayyub<sup>4</sup>, Matthias Mann<sup>5</sup>, David Garrick<sup>4,6,7</sup> and Douglas Vernimmen<sup>1,7</sup>**

<sup>1</sup> The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, United Kingdom.

<sup>2</sup> Institute of Molecular Biology (IMB), 55128 Mainz, Germany

<sup>3</sup> Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, London SE1 9RT, UK

<sup>4</sup> MRC Molecular Haematology Unit, Weatherall Institute for Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, United Kingdom.

<sup>5</sup> Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany.

<sup>6</sup> Current address : INSERM UMRS-976, Institut de Recherche Saint Louis, Université de Paris, 75010 Paris, France.

<sup>7</sup> These authors contributed equally

Correspondence:

[david.garrick@inserm.fr](mailto:david.garrick@inserm.fr), [douglas.vernimmen@roslin.ed.ac.uk](mailto:douglas.vernimmen@roslin.ed.ac.uk)

**Abstract word count: 123**

**Characters (including space): 36,848**

**Words count: 5,499**

**Figures: 6**

**Table: 1**

**Suppl Figures: 7**

**Suppl Tables: 6**

**References: 57**

**Running title:** MAZ regulates erythropoiesis

**Key words:** Erythropoiesis, Globin, Gene regulation, MAZ

## ABSTRACT

Erythropoiesis requires a combination of ubiquitous and tissue-specific transcription factors. Here, through DNA affinity purification followed by mass spectrometry, we have identified the widely expressed protein MAZ (Myc-associated zinc finger) as a transcription factor that binds to the promoter of the erythroid-specific human  $\alpha$ -globin gene. Genome-wide mapping in primary human erythroid cells revealed that MAZ also occupies active promoters as well as GATA1-bound enhancer elements of key erythroid genes. Consistent with an important role during erythropoiesis, knockdown of MAZ in primary human erythroid cells impairs erythroid differentiation, and genetic variants in the *MAZ* locus are associated with clinically important human erythroid traits. Taken together, these findings reveal the Zinc-finger transcription factor MAZ to be a previously unrecognised regulator of the erythroid differentiation program.

# INTRODUCTION

Over the last three decades, studies of the  $\alpha$ - and  $\beta$ -globin gene clusters have contributed to our understanding of some of the fundamental principles of mammalian gene regulation, including RNA stability, termination of transcription, and the identification of remote regulatory elements<sup>1</sup>. A complex network of DNA sequence elements, chromatin accessibility, histone modifications, and transcription factor (TF) occupancy orchestrates expression of globin genes, which are exclusively expressed in erythroid cells<sup>2</sup>. Both proximal regulatory regions (promoters) and distal regulatory elements (enhancers) are required for the initiation of  $\alpha$ -globin (*HBA*) and  $\beta$ -globin (*HBB*) gene expression in erythroid cells. When active, these regulatory elements are characterised by the presence of DNase I hypersensitivity sites (DHS) and active histone modifications. A small group of lineage-restricted TFs including GATA binding protein 1 (GATA1), T cell acute lymphocytic leukemia 1 protein (TAL1), and Erythroid Krüppel-like factor (EKLF; henceforth referred to as KLF1) act as erythroid ‘master regulators’<sup>3</sup>, by binding to both promoters and enhancers of the globin genes as well as to other genes important for erythropoiesis, thereby significantly contributing to their expression [reviewed in<sup>2,4,5</sup>].

Genetic and biochemical dissection of gene regulation programs has revealed that transcription of cell type-specific genes is usually achieved through synergy and cooperation between ubiquitously expressed TFs and tissue- or developmental stage-specific TFs<sup>6,7</sup>. Although this concept is well established, the mechanisms of interactions and relationships between tissue-specific and ubiquitous TFs are generally lacking. Moreover, the broad range of phenotypic abnormalities associated with mutations/alterations in ubiquitously expressed regulators preclude dissecting their roles in specific processes and determining their contribution to disease states.



Here, we have carried out an unbiased screen for proteins that bind to adjacent GC-rich motifs in the promoter of the duplicated  $\alpha$ -globin genes (HBA2 and HBA1) in human erythroid cells, combining electrophoretic mobility shift assays (EMSAs), DNA affinity purification, and mass spectrometry. This screen identified Myc-Associated Zinc-finger protein (MAZ) as a direct binder of the promoter region of the human  $\alpha$ -globin gene *in vitro* and *in vivo*. Knockdown of MAZ in primary human erythroid cells led to both reduction of  $\alpha$ -globin expression and a block of erythroid differentiation. Moreover, variants in the MAZ gene itself and its promoter region are associated with clinically important human erythroid traits. By ChIP-seq in primary human erythroblasts we showed that MAZ is enriched at transcription start sites (TSSs) of transcriptionally active genes and distal regulatory elements where it recognises a canonical G<sub>3</sub>(C/A)G<sub>4</sub> binding motif. Erythroid-specific MAZ signal is enriched at promoters and enhancers of genes associated with erythropoietic disorders. We found that MAZ erythroid-specific binding sites frequently colocalize with GATA1, a key regulator of erythropoiesis, particularly at enhancer elements, suggesting functional synergy between these two transcription factors. Together our findings have identified MAZ as an important regulator of the erythroid differentiation program.

## RESULTS

### **A novel DNA-protein complex occupies the distal promoter region of the human HBA genes.**

Transcriptional activation of the HBA genes during erythroid differentiation is associated with localised relaxation of chromatin structure, which is observed experimentally as a DHS immediately upstream (promoter region) of the adult HBA

genes (chr16: 172k-176k, Suppl. Figure 1A) specifically in erythroid cells<sup>8</sup>. In order to map chromatin accessibility within this erythroid-specific DHS at a higher resolution than provided by DNase- and ATAC-seq approaches, intact nuclei isolated from cells expressing  $\alpha$ -globin (the erythroid cell line K562) and not expressing  $\alpha$ -globin (EBV-transformed B lymphoblast cell line) were digested with low concentrations of selected restriction enzymes. The extent of digestion reflects accessibility of the specific recognition sites and was assessed by Southern blotting (Suppl. Figure 1B). This assay revealed that the region of erythroid-specific sensitivity extends from -220 bp (*FokI* site) to +35 bp (*NcoI* site) relative to the TSS of the HBA genes (Suppl. Figure 1B).

In order to characterise protein binding across this hypersensitive region, we carried out electrophoretic mobility shift assays (EMSAs) using a series of overlapping oligonucleotide probes to compare *in vitro* binding proteins in nuclear extracts prepared from the same erythroid (K562) and non-erythroid (EBV) cells. Oligonucleotide probes were specifically placed at positions suggestive of protein binding by *in vivo* dimethyl-sulfate (DMS) footprinting assays reported previously (Figure 1A, Table 1)<sup>9</sup>.

We observed band shifts for five out of seven probes used (Figure 1B). Previous motif analysis had suggested potential binding motifs for SP1 (-121/-116, -61/-56), CBF (-72/-65), Nuclear Factor I (NF1) (-85/-72), and  $\alpha$ -inverted repeat protein ( $\alpha$ -IRP) (-51/-42)<sup>9</sup> (Table 1). Gel supershift assays, in which nuclear extracts were pre-incubated with anti-CBF or anti-NF1 antibodies, affect the retarded bands at probes 7/8 and 9/10 respectively, confirming that these sites are indeed bound by CBF and NF1 (Figure 1B, lanes 13 and 17, respectively). Interestingly, complex binding patterns comprised of four shifted bands were observed at two neighbouring GC boxes situated at -100/-

83 (probe 11/12) and -128/-111 (probe 13/14). These GC boxes have a highly similar sequence (Suppl. Figure 2A) and were able to cross-compete with each other in competition assays (Suppl. Figure 2B), suggesting that they are likely to be bound by the same proteins. Interestingly, while the most heavily retarded bands (labelled a, b and c in Figure 1B) were detected in both K562 and EBV nuclear extracts, the faster migrating band (d) was observed only in K562 cells. This factor bound more strongly to probe 13/14 than to 11/12 (Figure 1B, compare lanes 20 and 23), a finding that was confirmed in competition assays (Suppl. Figure 2B). Gel shifts carried out with nuclear extracts from other cell types revealed that band (d) was the most abundant complex in primary human erythroblasts but was either not detected or weak in a range of non-erythroid cells types (HT-29, SW13, HepG2, and HeLa) (Figure 1C).

The probe 13/14 (covering -128/-111 region) contains a predicted binding motif for the Krüppel-like zinc-finger transcription factor Sp1<sup>9</sup> (Table1). We found that the binding of all four proteins to probe 13/14 was sensitive to the presence of the zinc-chelating agent EDTA (Suppl. Figure 3A), consistent with Zn-finger-dependent binding. In contrast, binding of the non-Zn-finger protein CBF to probe 7/8 was not affected by EDTA (Suppl. Figure 3A). To identify the protein responsible for these bands on probe 13/14, we carried out gel supershift reactions using antibodies against candidate Zn-finger proteins (Sp1, Sp3, Krüppel-like factor 1 (EKLF), and Krüppel-like factor 3 (BKLF)). These experiments revealed bands (a), (b) and (c) to be Sp1, Sp3 and KLF3, respectively (Figure 1D). These findings are consistent with the detection of proteins (a), (b) and (c) in both erythroid and non-erythroid nuclear extracts (Figure 1C), since Sp1, Sp3 and BKLF are all widely-expressed TFs and have previously been shown to regulate  $\alpha$ -globin expression<sup>10,11</sup>. In contrast, the binding of protein (d) to probe 13/14 was not strongly affected by any of the antibodies tested. Taken together, these

experiments have revealed an erythroid-enriched complex at neighbouring sites in the -128/-90 region upstream of HBA genes.

### **A mass spectrometry-based screen identifies MYC-associated Zn-finger protein MAZ as a factor binding at the $\alpha$ -globin promoter**

In order to identify the factor responsible for band (d), we carried out a protein affinity purification screen, using probe 13/14 as the bait (Figure 2A). As a negative control, we designed a mutant version of probe 13/14 (M3) in which a central guanidine was changed to thymidine, preventing formation of band (d) while also depleting binding of Sp1, Sp3 and BKLF (Suppl. Figure 3B). To carry out this screen, K562 nuclear extracts were incubated with desthiobiotin-modified DNA probes, protein-DNA complexes were purified using streptavidin beads, and eluted proteins were subjected to mass spectrometry (Figure 2A)<sup>12</sup>. Of the 390 proteins identified (Supplemental Table 1), one of the most interesting candidates to emerge was the Myc-associated Zn-finger protein MAZ, a C2H2-type Zinc finger protein which has been previously shown to bind *in vitro* to a G-rich consensus motif (G<sub>3</sub>AG<sub>3</sub>)<sup>13-15</sup> (Figure 2B).

In order to investigate whether MAZ is indeed the protein responsible for band (d), we performed siRNA-mediated knockdown (KD) of MAZ in K562 cells. Depletion of MAZ protein was associated with reduction of band (d) in the gel shift assays (Figure 2C). Conversely, expression of exogenous MAZ protein in COS7 cells was associated with strong enrichment of band (d) (Figure 2D). Binding of the over-expressed MAZ protein resulted in a depletion of the Sp3 gel shift, indicating competition for binding under these *in vitro* conditions. Altogether, these experiments confirm that MAZ is indeed the protein responsible for band (d) and indicate that MAZ binds to neighbouring GC boxes at -100/-83 and -128/-111 within the human  $\alpha$ -globin promoter *in vitro*.

In order to confirm binding by MAZ at the active HBA promoter *in vivo*, we carried out ChIP-qPCR experiments using a previously described series of qPCR amplicons throughout the HBA locus<sup>10</sup>. Whereas binding was absent or low in non-expressing cells (EBV lymphoblast), strong enrichment of MAZ was observed at the HBA promoter and gene body in cells expressing HBA, with the highest enrichment observed in primary erythroid cells (Figure 2E). In these primary cells, MAZ was also enriched at the important distal regulatory element of the  $\alpha$ -globin locus (MCS-R2). We further investigated the dynamics of MAZ recruitment to the  $\alpha$ -globin genes during erythroid maturation in primary erythroblast differentiation cultures (Suppl. Figure 4A-C). During differentiation, expression of the adult globin genes HBB and HBA is strongly upregulated, peaking in intermediate and late stage erythroblasts (Suppl. Figure 4D, see also <sup>16</sup>). Concomitantly with the strong upregulation of  $\alpha$ -globin expression, MAZ was dynamically recruited to the HBA promoter and genes in these cultures (Suppl. Figure 4E). Altogether, our results indicate that MAZ binds to the active human HBA locus both *in vitro* and *in vivo*.

### **MAZ is required for erythroid differentiation and is associated with clinical erythroid-related traits**

To investigate the functional significance of MAZ during erythroid differentiation, we used shRNA to deplete MAZ expression in differentiating primary human erythroid cultures. FACS analysis revealed an impaired differentiation of primary erythroid cultures infected with the MAZ shRNA lentivirus, with cells accumulating at the CFUe and pro-erythroblast (CD71<sup>+</sup>GPA<sup>lo</sup>) stages, and failing to differentiate towards intermediate erythroblasts (CD71<sup>+</sup>GPA<sup>+</sup>) (Figure 3A, B). This impaired differentiation could also be observed as a deficiency of haemoglobinisation (Figure 3C). Knockdown of MAZ expression was further associated with downregulation of expression of  $\alpha$ -

globin (Figure 3D). While this downregulation of  $\alpha$ -globin is likely to also arise indirectly due to the impaired differentiation, expression of  $\alpha$ -globin was more affected than  $\beta$ -globin (where no binding of MAZ was observed in erythroid cells, see below), consistent with MAZ playing a direct regulatory role at the  $\alpha$ -globin locus.

Given these findings suggesting a previously unknown for MAZ during erythroid differentiation, we further investigated whether human genetic variants in the *MAZ* gene are associated with clinically important erythroid traits. We analysed the GeneAtlas database that reports associations between 778 traits and millions of DNA variants <sup>17</sup>. Out of 25 erythroid-related traits in the GeneAtlas (Suppl. Table 2), eight traits (32%) were associated with three variants in the *MAZ* gene or promoter region (rs11559000, rs572982482, rs72798129) with *p*-value  $10^{-2}$ - $10^{-5}$  (Figure 3E). Taken together, these findings indicate that MAZ is an important factor in the erythroid differentiation program and suggest that this locus contributes to clinically relevant erythroid traits.

### **MAZ occupies TSS and binds directly to DNA through G<sub>3</sub>(C/A)G<sub>4</sub> motif**

Following from our findings at the HBA locus, we expanded our analysis by investigating binding of MAZ genome-wide in primary human erythroid cells by carrying out ChIP-seq (Suppl. Figure 5A-B). As previously observed by ChIP-qPCR (Figure 2E), MAZ was strongly enriched at both HBA2/1 promoters as well as, to a weaker extent, at the MCS-2 remote regulatory element (Figure 4A, top). In contrast, very low binding of MAZ was detected at the promoter and regulatory elements (locus control regions; LCR) of the  $\beta$ -globin gene cluster (Figure 4A, bottom). Genome-wide, we identified 10 088 MAZ binding sites in primary erythroid cells (Suppl. Figure 5C).

While the majority of MAZ binding sites (65%) were located at promoter regions (Figure 4B), the average MAZ enrichment was comparable for peaks in the promoter regions, intergenic, and genic regions (Suppl. Figure 5D). Overall, MAZ was present at least at one TSS of 28% (5466/19646) of protein-coding genes. Comparison with published ChIP- and ATAC-seq datasets from erythroid cells<sup>8,18-20</sup> (Suppl. Table 6) revealed that MAZ binding sites are located in regions of high chromatin accessibility (as detected by ATAC-seq) that are enriched for activating histone modifications (H3K4me3, H3K27ac) and Pol II but depleted of the repressive histone mark H3K27me3 (Figure 4C). Taken together, these results indicate that MAZ is enriched around TSS of a large proportion of transcriptionally active genes in erythroid cells.

In order to identify motifs contributing directly to MAZ binding, we used the MEME software suite<sup>21</sup> to discover *de novo* enriched sequence motifs among a training set consisting of the 500 highest ranked MAZ peaks<sup>22</sup>. Overall, six enriched motifs were detected in the MAZ erythroblast training dataset (E-value <0.01) (Figure 4D). The most significant motif found, G<sub>3</sub>(C/A)G<sub>4</sub>, is contained within the 13/14 probe derived from the HBA promoter used to identify MAZ in the MS screen, and was similar to the published canonical MAZ motif G<sub>3</sub>AG<sub>3</sub><sup>23</sup>. This canonical motif was present in 92% of all MAZ peaks (see methods and Figure 4D), and was the only enriched motif with a narrow unimodal central enrichment in the peaks and a large maximum site probability (Figure 4E), suggesting that the vast majority of MAZ genomic binding in erythroblasts is due to direct DNA binding of MAZ to this canonical DNA sequence motif. In agreement with these observations, mutations within this core G<sub>3</sub>CG<sub>4</sub> motif present in the 13/14 probe prevented MAZ binding *in vitro*, while mutations outside this core motif had little effect (Suppl. Figure 6). Furthermore, the 11/12 probe, which exhibits lower affinity for MAZ (Figure 1B and Sup Fig. 2C), contains a less-perfect version of this

core motif than the 13/14 probe (Suppl. Figure 6). The second most abundant motif detected in our MAZ training set was a GGAGGA-containing motif. This motif was present in 31% of MAZ peaks, but did not exhibit central localisation, suggesting it may contribute to MAZ binding in a cooperative manner. The enrichment of this motif is consistent with the previous reports demonstrating binding by MAZ to GGA repeats with a high propensity to form G4-quadruplexes <sup>24-27</sup>.

### **Erythroid-specific MAZ binding sites are associated with the promoters of key erythropoiesis genes**

To gain further insight into the specific role of MAZ during erythroid differentiation, we compared the genome-wide profile of MAZ binding in primary erythroblasts with MAZ ChIP-seq profiles from five human non-erythroid cell lines (HepG2, GM12848, MCF-7, IMR90, A549) generated by the ENCODE consortium <sup>28,29</sup> (Suppl. Table 6). Comparison of MAZ peaks identified in these six cell types revealed that, as expected for a housekeeping transcription factor, a large proportion of MAZ peaks are shared between at least two different cell types (so called “common” peaks, (n=8308) (Figure 5A). However, between 8 and 40% of the MAZ peaks observed in a given cell line were not observed in any other cell type (Figure 5A, Suppl Table 3), suggesting that as well as regulating housekeeping functions, MAZ also plays an important role in the control of cell-type restricted gene expression programs. Consistent with this, Pearson correlation coefficients for these datasets indicated that MAZ exhibited globally distinct binding profiles in the different cell types (correlation coefficient < 0.85 for all pairwise comparisons) (Figure 5B). In particular, 18% of MAZ peaks observed in erythroid cells were not shared with other cell types (1780 peaks, termed “erythroid-specific peaks”). Comparison with gene expression profiles <sup>20</sup> revealed that MAZ binds 41% (218/528)



of promoters of genes with erythroid-specific expression. Gene ontology analysis revealed that genes with a TSS bound by MAZ in an erythroid-specific manner were significantly associated with erythroid differentiation and blood haemostasis functions (Figure 5C) as well as phenotypes associated with haematological diseases (Figure 5D). We also investigated MAZ binding at genomic loci which have previously been linked to clinical erythroid phenotypes in GWAS studies. Out of 31 genes associated with erythropoiesis<sup>30</sup>, MAZ peaks were present on the promoters of 21 of them (Suppl. Table 4). Interestingly, among these genes with promoters bound by MAZ in an erythroid-specific manner were the TFs GATA1 and KLF1, master regulators of erythropoiesis (Suppl. Table 4 and Suppl. Figure 7). Taken together, these findings indicate that MAZ binds to the promoters of genes with key roles in erythroid differentiation and homeostasis.

### **Erythroid-specific MAZ signal is enriched on GATA1-bound enhancers**

Interestingly, the common and erythroid-specific peaks of MAZ displayed distinct localisation relative to genomic features (Figure 6A). While the majority (71%) of common peaks were localised at promoter regions, only 24% of the MAZ erythroid specific peaks were located at promoters. In contrast to common MAZ peaks, erythroid-specific MAZ peaks were enriched at intergenic (24%) and intronic (40%) sites, suggesting that MAZ may play a particularly important role at erythroid-specific distal regulatory elements. Comparison with our previously published catalogue of erythroid enhancers<sup>20</sup> confirmed that 27% of the erythroid-specific MAZ peaks coincided with known enhancer elements, compared with only 7% of the “common” peaks. Consistent with this observation, we observed enrichment of the enhancer-associated histone modification H3K4me1 on erythroid-specific non-TSS MAZ peaks,

but not on common non-TSS MAZ peaks, with H3K4me3 following the reverse trend (Figure 6B). Taken together, these findings suggest that erythroid specific binding of MAZ is particularly important at distal enhancer elements.

Ubiquitous and lineage specific factors often work together at proximal and distal regulatory sequences. As such, we explored the association between MAZ and the erythroid master regulator GATA1 in primary erythroid cells. Overall, 48% (854/1778) of erythroid-specific MAZ binding sites overlapped with GATA1 binding sites, consistent with a functional cooperative interaction between these two factors. Interestingly, this co-association between MAZ and GATA was particularly prominent within erythroid enhancer regions, with GATA1 signal being particularly elevated at erythroid MAZ peaks within enhancers and conversely, MAZ signal being stronger at GATA1 enhancer peaks (Figure 6C). Taken together, these findings are consistent with a particular functional cooperativity between MAZ and GATA1 at erythroid enhancer elements.

## DISCUSSION

In this study, we have identified the ubiquitously-expressed Zn-finger protein MAZ as a factor binding to neighbouring GC-rich sites within the human  $\alpha$ -globin promoters. ChIP and ChIP-seq experiments subsequently confirmed *in vivo* binding of MAZ to the active HBA genes as well as to promoters and enhancers of other key genes within the erythroid differentiation pathway and to genomic loci linked to clinical erythroid phenotypes. Moreover, erythroid-specific binding of MAZ was enriched at distal regulatory enhancer elements, where it frequently co-associated with the erythroid master-regulator GATA1 and knockdown of MAZ impaired erythroid differentiation in

*ex vivo* primary cultures. We further showed that genetic variants within the MAZ locus were associated with approximately one third of human phenotypic erythroid traits. Taken together, these findings reveal a previously unrecognised role for MAZ in the erythroid differentiation program.

MAZ has not previously been shown to play an important role in the erythroid lineage. Germline deletion of MAZ in mice results in perinatal lethality<sup>31</sup>, as might be expected for a ubiquitously expressed TF. MAZ knockdown in a primary erythroid differentiation culture system strongly affects HBA expression whereas expression of HBB was less affected. This is consistent with our ChIP-seq data showing binding of MAZ to the HBA promoters and distal regulatory elements and support a direct role for MAZ at this locus. Since our experiments indicate impaired differentiation of primary erythroid cultures infected with the MAZ shRNA lentivirus, inducible knockdown or knockout approaches will be needed to precisely determine the role of MAZ at specific stages of erythroid maturation.

Through integration of MAZ binding profiles in different cell types, we observed both common and erythroid-specific MAZ binding sites. Erythroid-specific MAZ binding is observed at the promoters of *GATA1* and *KLF1*, encoding master-regulators of the erythroid lineage. MAZ binding is also enriched at distal regulatory elements, which are considered to be primary determinants of tissue-specific gene expression programs<sup>18,32</sup>. Importantly, it has been shown that the activity of erythroid enhancer elements cannot be predicted based on the binding of the master regulators of erythroid differentiation, GATA1 and TAL1, alone<sup>18</sup>. In contrast, combinatorial co-occupancy of enhancers by both lineage-specific and ubiquitously expressed transcription factors is a more reliable indicator of enhancer activity and cell-specific expression<sup>18,32,33</sup>. Our findings suggest that binding of MAZ

together with lineage-restricted factors such as GATA1 might be an important step in the activation or maintenance of many erythroid enhancer elements.

The vast majority (92%) of both shared and erythroid-specific MAZ peaks contain the canonical G<sub>3</sub>CG<sub>4</sub> motif, indicating that MAZ binds DNA directly in both cases. Consistent with previous reports<sup>24-27</sup>, our analysis of MAZ whole genome binding also indicated that a likely G4-forming motif (GGAGGA) contributes to MAZ binding, although our data suggests that this motif makes only a secondary contribution to MAZ binding affinity.

At present, it is unclear how binding of MAZ to its target sites is regulated. In particular, while MAZ is a ubiquitously expressed protein whose levels do not change dramatically during erythroid differentiation (data not shown), we observed differential *in vitro* binding by MAZ to the 13/14 probe in gel shift assays, as well as a highly dynamic recruitment of MAZ to the HBA locus in primary erythroid cultures, suggesting that its binding is not primarily regulated at the level of expression. Western blot with  $\alpha$ -MAZ antibody detected several bands in K562 cells, which were sensitive to MAZ shRNA (Figure 2C), suggesting the possibility that MAZ is subject to post-translational modifications in these cells. Interestingly, it has been shown in other cell types that phosphorylation at Ser480 (by casein kinase II)<sup>34</sup> or at Ser187 and Thr386 (by protein kinase A), increases the DNA-binding and transcriptional activation properties of MAZ<sup>35</sup>. MAZ can also be phosphorylated at Thr71 downstream of the Mitogen-activated protein (MAP) kinase signalling pathway, and this phosphorylation is critical for activation of MAZ in response to inflammatory cytokines<sup>36</sup>. In the future it will be important to characterise how post-translational modifications regulate MAZ activity at the HBA locus and during erythroid differentiation, as well as the upstream signalling pathways involved. Interestingly, our data demonstrate that the sites bound by MAZ

within the HBA promoter can also be bound *in vitro* by other KLFs, including SP1, SP3 and BKLF, suggesting the possibility of a crosstalk between these factors in regulating the expression of  $\alpha$ -globin. Indeed, complex cross-regulatory interactions involving MAZ and Sp1 have also been reported at other promoters<sup>37,38</sup>. We also found that these factors compete for the same binding site in our *in vitro* conditions. It is of interest that the expression of SP1 decreases markedly during erythroid differentiation<sup>39</sup> and that this decline correlates well with the dynamic recruitment of MAZ to the HBA promoters that is observed in primary erythroid differentiation cultures. These observations suggest that MAZ may bind to the HBA promoters in a commensurate manner when SP1 levels decline during erythroid differentiation, and that this replacement of SP1 by MAZ during the late stages of erythroid differentiation correlates with the maximal activation of the HBA genes. In the future it will be important to explore this potential crosstalk between SP1 and MAZ at the HBA promoters and how it plays out at the genome-wide level during erythroid differentiation.

Overall, this study serves as an example of how a precise molecular characterisation of protein binding at a single regulatory element can be combined with an unbiased mass-spectrometry screen to identify new trans-acting regulators. This approach has revealed the Zn-finger transcription factor MAZ as a previously-unrecognised regulator of the erythroid differentiation program, which may be important for human erythroid phenotypic traits.

# **METHODS**

## **Cell lines and primary erythroid cells**

Primary erythroid cells and newly generated EBV-infected B-lymphoblasts were obtained as previously described <sup>39</sup>. Cell lines (K562, HT29, HeLa, SW13 and COS7) were cultured in RPMI 1640 supplemented with 10% FBS. For siRNA-mediated knockdown of MAZ, K562 cells were transfected with MAZ SMARTPool siRNA or Non-targeting control pool siRNA (Dharmacon) using Lipofectamine according to the manufacturer's instructions, and cells were harvested after 4 days. For overexpression of MAZ, COS7 cells were transfected with the plasmid pcDSAF-1, which expresses full length MAZ (SAF-1) cDNA under the control of the CMV promoter <sup>40</sup>. Cells were transfected using Lipofectamine according to the manufacturer's instructions and were harvested after 30 h.

## **Electrophoretic mobility-shift assay (EMSA)**

Nuclear extracts were performed as described <sup>41</sup> except that nuclei were incubated in buffer C for 30 min. Protein concentration was determined using the Qubit Protein Assay Kit (ThermoFisher). Oligonucleotide probes (Table 1) were designed to have an additional 5'GG overhang on each end after annealing and were labelled by filling in with the large Klenow fragment of DNA Polymerase I in the presence of [ $\alpha$ -<sup>32</sup>P]-dCTP as described by the manufacturer (New England Biolabs). For gel shift reactions, 5  $\mu$ g of nuclear extract was incubated with 1 ng radiolabelled probe (>10,000 cpm) in buffer [10 mM HEPES (pH 7.8), 50 mM potassium glutamate, 5 mM MgCl<sub>2</sub>, 1 mM EDTA, 1 mM dithiothreitol, 0.5  $\mu$ g of poly(dI-dC), 10  $\mu$ g of bovine serum albumin, 1 mM ZnSO<sub>4</sub>, 5% glycerol] for 30 min on ice. For competition experiments, unlabelled competitor oligonucleotides were added to the binding reactions at 100 X molar excess. For gel

supershift experiments, antibodies were incubated with nuclear extract for 20 min on ice prior to the addition of the radiolabelled probe. Antibodies used in supershift experiments were as described in previous ChIP studies <sup>10</sup> . Following binding, samples were subjected to electrophoresis at 4°C for 2.5 h at 12 V/cm on a native polyacrylamide gel (6% [19:1] *bis*:acrylamide in 0.5 X Tris-borate-EDTA). The gels were dried and analysed using a Storm 860 Molecular Imager (Molecular Dynamics Ltd).

### **shRNA knockdown in primary erythroid cultures**

Lentiviral constructs (pLKO.1) expressing MAZ-specific (clone TRCN0000015343) or non-targeting scrambled shRNAs were obtained from Dharmacon. These constructs express the shRNA from the human U6 promoter and the Puro<sup>r</sup> gene from the hPGK promoter. Lentiviral preparations were prepared using standard techniques by co-transfection of 293T cells with the lentiviral construct together with pMDLg/pRRE and pMD2.G using Fugene (Roche). Culture supernatants were harvested at 72 h after transfection and concentrated by ultracentrifugation using standard techniques. For lentiviral-mediated knockdown experiments, primary erythroid differentiation cultures were carried out essentially as described in Leberbauer et al., <sup>42</sup> with the following modifications. Peripheral blood mononuclear cells were cultured for 8 days in erythroblast expansion medium (StemSpan™ SFEM medium (Stem Cell Technologies) supplemented with Epo (2 U/mL), IGF-1 (40 ng/mL), SCF (100 ng/mL) (all cytokines from Peprotech), dexamethasone (1 µM, Sigma) and cholesterol-rich lipids (40 µg/mL, Sigma)). Cells were infected with concentrated lentivirus for 24 h in wells pre-coated with Retronectin (0.2 mg/ml; Takara) in expansion medium supplemented with protamine sulfate (10 µg/ml). The cells were seeded in fresh complete medium supplemented with cyclosporine A (1 mg/ml) for 24 h prior to selection with puromycin (1 µg/ml) for 48 h. After removing Puromycin, cells were

cultured for further 3 days in expansion medium containing cyclosporine prior to harvesting for FACS and RNA (total 15 days in culture).

## Mass Spectrometry

For proteomics analysis, the DNA baits corresponding to oligonucleotide 13/14 WT and 13/14 M3 with a TT/AA-overhang were annealed, phosphorylated, ligated and purified as previously described<sup>12</sup>. The desthiobiotinylated oligonucleotides were coupled to Streptavidin Dynabeads MyOne C1 (Life Technologies) for 60 min at room temperature and excess oligonucleotides removed by washing. The oligonucleotide coupled streptavidin beads were incubated with 250 µg of K562 nuclear extract diluted in PBB buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 0.5 % Igepal CA630, Complete Protease Inhibitor without EDTA (Roche)) for 2 hours at 4°C under slight agitation. Non-bound proteins were removed by washing three times with PBB buffer and the bound DNA-protein complexes liberated from the streptavidin beads with 200 µl of a 16 mM biotin/50 mM ABC (pH 8.0) solution. The complexes were precipitated with pure ethanol overnight at room temperature. The pellet was resuspended in 50 µl 8M urea and first digested with LysC (Wako) for 3 hours and subsequently diluted to 250 µl with 50mM ABC buffer and digested with 200 ng trypsin overnight, both at room temperature. The tryptic peptides were loaded onto an SCX StageTip<sup>43</sup> and stored until the MS measurement.

The tryptic peptides were fractionated on an in-house packed 25 cm microcapillary column (Reprosil 3.0 µm; Dr. Maisch GmbH) in a 125 minute gradient from 5 to 60 % acetonitrile. The MS measurement was performed on an LTQ Orbitrap XL (Thermo) with a top5 DDA method using CID fragmentation. The MS data was processed with MaxQuant<sup>44</sup> version 1.0.12.5 using the human Uniprot database.



## Expression analysis

RNA was extracted with Tri reagent (Sigma) according to the manufacturer's instructions. RNAs were DNaseI treated (Ambion) and cDNA was generated with SuperScript III (Invitrogen) as previously described <sup>45</sup>.

## ChIP-qPCR and ChIP-Seq assay

ChIP was performed as previously described <sup>46</sup>. Briefly, chromatin was first cross-linked with ethylene glycol bis(succinimidyl succinate) (EGS) in PBS at a final concentration of 2 mM for 60 min at RT. Formaldehyde (CH<sub>2</sub>O) was then added at a final concentration of 1 % for 15 min at RT and samples were sonicated over 20 min (10 x 30 s pulses) at 4°C to cleave genomic DNA (Bioruptor, Diagenode). Anti-MAZ antibody was from Bethyl Laboratories (A301-652A). Real-Time PCR was performed using primers and probes (5'FAM-3'TAMRA) for the human  $\alpha$ -globin locus described previously <sup>10</sup>. Each ChIP was performed as two independent experiments and quality was assessed by qPCR. The ChIP-seq libraries were prepared using the New England Biolabs NEBNext® ChIP-Seq Library Prep Reagent Set for Illumina according to the manufacturer's protocol and starting with between 6 and 40 ng of captured DNA. All libraries received 12 cycles of PCR amplification in the final step before PCR clean-up using Ampure beads. Sequencing was carried out on the HiSeq 2500 using Illumina HiSeq Rapid Cluster Kit v2 - Paired-End for 100 cycles. RTA software version used was Illumina RTA 1.17.20 and the Pipeline software version was bcl2fastq-1.8.3.

## ChIP-Seq analysis

**Alignment.** Reads were aligned using the hisat2 aligner <sup>47</sup> version 2.0.3 with the --no-spliced-alignment option, but otherwise default parameters, to a splicing-unaware index of the human GRCh38 genome.

**The strength of enrichment in the IP** was assessed plotting “fingerprints” using deeptools <sup>48</sup> and by calculating normalized strand cross-correlation coefficient (NSC) and relative strand cross-correlation coefficient (RSC) metrics using Phantompeakqualtools <sup>49</sup>.

**Normalisation of the ChIP-Seq signal.** Reads of the ChIP-Seq samples were first normalised to the input and then scaled to 1x sequencing depth using deepTools v. 3.1.3 <sup>50</sup>.

**Peak calling.** Peak calling was performed using MACS2 2.1.1 with the minimum FDR (q-value) cutoff of 0.01 <sup>51</sup>. The top 75% fold enriched peaks were selected for further analysis. For ENCODE datasets optimal IDR thresholded peaks provided by the ENCODE consortium have been used.

**Peak annotation** was made with UCSC RefSeq gene annotation (GRCh38 genome version) using HOMER suite 4.8 <sup>52</sup>. For these analyses we define promoters as regions located within a distance of -3000 bp upstream and 100 bp downstream of the TSS.

**Coverage analysis and ChIP heatmap plots and profiles** were performed using DeepTools suites. General genome arithmetics was performed using BEDTools v. 2.27.1 <sup>53</sup>. The set of erythroid-specific and housekeeping genes and erythroid enhancers was based on <sup>20</sup>.

**Correlation of the datasets.** A clustered heatmap of correlation coefficients of bigwig signal was computed using the Pearson method; the bam signal combined from the replicates was plotted over 10 kb bins with DeepTools.

**Defining erythroid-specific MAZ signal.** The peaks called on MAZ datasets were overlapped with the peaks from five MAZ ENCODE datasets with the minimum overlap

of 1 bp using BEDTools. The MAZ peaks without the overlap in any of the regions were assigned to “erythroid-specific peakset”.

**Motif Analysis.** To find the sequence motifs enriched in MAZ peaks, genomic sequences (from -50 bp to +50 bp around the centres of the top 500 MAZ peaks ranked on their q-values) were extracted (hg38) and used as input for MEME *de novo* motif discovery<sup>54</sup> with *E*-value cutoff 0.01 and motif size defined as 6-30 bp.

To calculate the presence of the motifs in a given peak, we used FIMO with the above defined motifs at a *p*-value threshold of  $10^{-4}$ <sup>55</sup>. Motif central enrichment analysis was performed using CentriMo<sup>22</sup> on regions from -250bp to +250 bp relative to the peak summits.

**Functional analysis of genes with erythroid-specific MAZ enrichment.** g:Profiler<sup>56</sup> was employed to conduct Gene Ontology (GO) Biological Function and HP (Human Phenotype) gene annotations. Fisher's exact test was used to retrieve significantly enriched GO terms for genes marked with erythroid-specific MAZ signal. Functional categories are defined as those containing at least five genes and a minimum enrichment score of 1.3 (*p*-value < 0.05). GeneAtlas database<sup>17</sup> was used for mining MAZ variants associated with erythroid-specific traits (*p*-value <  $10^{-2}$ ).

## **Data sources, protocols and analysis.**

Sources of ChIP-seq data are shown in Suppl. Table 6.

## **DECLARATIONS**

### **Authors' contributions**

D.G. and D.V. developed the hypothesis; F.B., M.L.H., V.S., J.A.S.S., H.A., and D.G. performed experiments and/or collected data; D.D., F.B., M.M. D.G. and D.V. analyzed

data; interpreted data; and D.D., D.G., and D.V. wrote the manuscript, which was revised and approved by all authors. All authors read and approved the final manuscript.

## **Acknowledgements**

We would like to thank Doug Higgs, Robert Beagrie, Michele Goodhardt and Philipp Voigt for critically reading the manuscript. High-throughput sequencing was provided by Edinburgh Genomics (<http://genomics.ed.ac.uk>).

## **Competing interests**

The authors declare that they have no competing interests.

## **Ethics approval and consent to participate**

Study protocols received ethical approval from Oxford University Ethical Review Panel. The data were analysed anonymously.

## **Funding**

This work was supported by a University of Edinburgh Chancellor's Fellowship to Douglas Vernimmen and by Institute Strategic Grant funding to the Roslin Institute from the BBSRC [BB/J004235/1] and [BB/P013732/1]. Darya Deen is supported by Roslin Institute core funding to Douglas Vernimmen. David Garrick was supported by the Medical Research Council (UK) and INSERM (France).

## FIGURE CAPTIONS

**Figure 1. An erythroid-enriched complex binds a distal element of the HBA promoter. (A)** Structure of the HBA core and distal promoter elements. The TSS is depicted as an angled arrow, and the locations of the EMSA probes are shown as grey bars. *In vivo* DMS footprints detected in K562 and erythroblasts<sup>9</sup> are represented by the black boxes. The nucleotide located 18 bp downstream of the TATA box (\*) differs between the HBA1 (C) and HBA2 (G) promoters. **(B)** EMSA and supershift assays using K562 and EBV nuclear extracts with the indicated probes. **(C)** The probe 13/14 binds an additional protein enriched in erythroid cells. **(D)** Sp/X-KLF-family antibodies cause retardation or disruption of the bands (a), (b) and (c), but not (d).

**Figure 2. Identification of MAZ by mass spectrometry. (A)** Schematic representation of the affinity purification screen. **(B)** The 10 proteins identified by mass spectrometry which were most enriched for binding to the wild type probe relative to the mutant. MAZ is highlighted in red. **(C)** and **(D)** EMSA showing that the intensity of band (d) is dependent on MAZ expression using knockdown (C) or overexpression (D) experiments. Knockdown experiments (siRNA) in K562 showed a strong reduction of MAZ protein levels and reduction of complex (d) in EMSA assays. Ectopic expression of MAZ (SAF1) in COS7 cells strongly enhanced complex (d) in EMSA assays. **(E)** MAZ is recruited to the active  $\alpha$ -globin promoter *in vivo*. Analysis of MAZ binding at the  $\alpha$ -globin locus in EBV-lymphoblasts, K562 cells and human primary erythroid cells by ChIP-qPCR. The y axis represents enrichment over the input DNA, normalised to a control sequence in the human 18S gene. The x axis indicates the Taqman probes used. The position of probes within the  $\alpha$ -globin cluster are indicated on the heading map. The  $\alpha$ -globin genes themselves are covered by three probes (Pr/Ex1, Ex2, Ex3). Error bars correspond to one SEM from two independent ChIPs.

**Figure 3. MAZ Knockdown impairs erythropoiesis. (A)** Flow cytometry analysis of CD71 and GPA expression at day 15 of representative primary erythroblast differentiation cultures infected with lentivirus expressing scramble shRNA or MAZ shRNA. In scramble shRNA cultures, cells progress towards a CD71<sup>+</sup>GPA<sup>+</sup> phenotype whereas shRNA against MAZ blocked differentiation at CD71 single positive cells. **(B)** Quantitation of flow cytometry staining as demonstrated in (A). Shown is the mean and standard deviation from two independent differentiation cultures. **(C)** Cell pellets from primary cultures at d15 of differentiation. Cultures with the MAZ shRNA are more pale, indicating less extensive hemoglobinisation. **(D)** Real-time RT-PCR analysis of expression of MAZ, HBA1, HBA2 and HBB after 15 d of primary erythroid differentiation cultures. Shown is the mean and standard deviation from two independent differentiation cultures. For each gene, expression (relative to the GAPDH) is normalized to the mean value observed with scramble shRNA. **(E)** Variants around the MAZ locus significantly associated with clinical erythroid traits (p-value <10<sup>-2</sup>).

**Figure 4. MAZ is enriched at active TSS and binds to DNA through a (G)<sub>3</sub>C(G)<sub>4</sub> consensus site. (A)** MAZ ChIP-seq enrichment profiles for the  $\alpha$ - and  $\beta$ -globin loci. **(B)** Genomic distribution of MAZ binding sites showing the number of peaks overlapping the category of genomic element indicated. **(C)** Heatmap plots of ATAC-seq, H3K4me3, H3K27ac, Pol II and H3K27me3 signal centred on MAZ peaks in erythroblasts cells (sorted according to decreasing MAZ ChIP signal). **(D)** Characteristics of DNA motifs significantly enriched (E<0.01) in MAZ peaks. **(E)** Motif localisation curves relative to the MAZ peak centres. The numbers of the enriched motifs correspond to the logos in (D).

**Figure 5. Analysis of erythroid-specific MAZ signal.** (A) Heatmap plots of MAZ ChIP-seq datasets from six cell types centered on MAZ common and cell-line-specific peaksets. Numbers on left indicate the common peaks (1), and peaks specific to erythroblasts (2), HepG2 (3), A549 (4), GM12878 (5), MCF-7 (6) and IMR90 (7) cells. (B) Pearson correlation analysis of MAZ ChIP-seq datasets from the six cell types. (C) Gene ontology (by biological function) of genes with erythroid-specific MAZ signal. (D) Human Phenotype Ontology of genes with erythroid-specific MAZ signal.

**Figure 6. Erythroid-specific MAZ signal is enriched on GATA1-bound enhancers.** (A) Genome distribution of MAZ common and erythroid-specific binding sites. (B) Average H3K4me1, H3K4me3, MAZ and ATAC-seq signal in primary erythroid cells plotted against MAZ erythroid-specific peakset (black) and MAZ common peakset (red) where signal from promoters have been excluded (non-TSS set). (C) Left panel: Heatmap plot of GATA1 ChIP-seq centred on MAZ erythroid peaks overlapping enhancers (top) and promoters (bottom). Right panel: Heatmap plot of MAZ ChIP-seq centred on GATA1 peaks overlapping enhancers (top) and promoters (bottom).

## REFERENCES

1. Vernimmen D. Globins, from Genes to Physiology and Diseases. *Blood Cells Mol Dis*. 2018;70:1.
2. Philipsen S, Hardison RC. Evolution of hemoglobin loci and their regulatory elements. *Blood Cells Mol Dis*. 2018;70:2-12.
3. Wontakal SN, Guo X, Smith C, et al. A core erythroid transcriptional network is repressed by a master regulator of myelo-lymphoid differentiation. *Proc Natl Acad Sci U S A*. 2012;109(10):3832-3837.
4. Perkins A, Xu X, Higgs DR, et al. Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood*. 2016;127(15):1856-1862.
5. Katsumura KR, DeVilbiss AW, Pope NJ, Johnson KD, Bresnick EH. Transcriptional mechanisms underlying hemoglobin synthesis. *Cold Spring Harb Perspect Med*. 2013;3(9):a015412.
6. Higgs DR, Vernimmen D, Wood B. Long-range regulation of alpha-globin gene expression. *Adv Genet*. 2008;61:143-173.
7. Vernimmen D. Uncovering enhancer functions using the alpha-globin locus. *PLoS Genet*. 2014;10(10):e1004668.
8. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016;48(10):1193-1203.
9. Rombel I, Hu KY, Zhang Q, Papayannopoulou T, Stamatoyannopoulos G, Shen CK. Transcriptional activation of human adult alpha-globin genes by hypersensitive site-40 enhancer: function of nuclear factor-binding motifs occupied in erythroid cells. *Proc Natl Acad Sci U S A*. 1995;92(14):6454-6458.
10. Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J*. 2007;26(8):2041-2051.
11. Funnell AP, Vernimmen D, Lim WF, et al. Differential regulation of the alpha-globin locus by Kruppel-like factor 3 in erythroid and non-erythroid cells. *BMC Mol Biol*. 2014;15(1):8.
12. Butter F, Davison L, Viturawong T, et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet*. 2012;8(9):e1002982.
13. Himeda CL, Ranish JA, Hauschka SD. Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. *Mol Cell Biol*. 2008;28(20):6521-6535.
14. Parks CL, Shenk T. Activation of the adenovirus major late promoter by transcription factors MAZ and Sp1. *J Virol*. 1997;71(12):9600-9607.
15. Ashfield R, Patel AJ, Bossone SA, et al. MAZ-dependent termination between closely spaced human complement genes. *EMBO J*. 1994;13(23):5656-5667.
16. Brown JM, Leach J, Reittie JE, et al. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol*. 2006;172(2):177-187.
17. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. 2018;50(11):1593-1599.
18. Xu J, Shao Z, Glass K, et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell*. 2012;23(4):796-811.



19. Huang J, Liu X, Li D, et al. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell*. 2016;36(1):9-23.
20. van de Lagemaat LN, Flenley M, Lynch MD, et al. CpG binding protein (CFP1) occupies open chromatin regions of active genes, including enhancers and non-CpG islands. *Epigenetics Chromatin*. 2018;11(1):59.
21. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. 2011;27(12):1696-1697.
22. Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res*. 2012;40(17):e128.
23. Song J, Ugai H, Ogawa K, et al. Two consecutive zinc fingers in Sp1 and in MAZ are essential for interactions with cis-elements. *J Biol Chem*. 2001;276(32):30429-30434.
24. Lew A, Rutter WJ, Kennedy GC. Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc Natl Acad Sci U S A*. 2000;97(23):12508-12512.
25. Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res*. 2008;36(6):1755-1769.
26. Cogoi S, Paramasivam M, Membrino A, Yokoyama KK, Xodo LE. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J Biol Chem*. 2010;285(29):22003-22016.
27. Membrino A, Cogoi S, Pedersen EB, Xodo LE. G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. *PLoS One*. 2011;6(9):e24421.
28. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.
29. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46(D1):D794-D801.
30. Tumburu L, Thein SL. Genetic control of erythropoiesis. *Curr Opin Hematol*. 2017;24(3):173-182.
31. Haller M, Au J, O'Neill M, Lamb DJ. 16p11.2 transcription factor MAZ is a dosage-sensitive regulator of genitourinary development. *Proc Natl Acad Sci U S A*. 2018;115(8):E1849-E1858.
32. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108-112.
33. Stadhouders R, Cico A, Stephen T, et al. Control of developmentally primed erythroid genes by combinatorial co-repressor actions. *Nat Commun*. 2015;6:8893.
34. Tsutsui H, Geltinger C, Murata T, et al. The DNA-binding and transcriptional activities of MAZ, a myc-associated zinc finger protein, are regulated by casein kinase II. *Biochem Biophys Res Commun*. 1999;262(1):198-205.
35. Ray A, Ray P, Guthrie N, Shakya A, Kumar D, Ray BK. Protein kinase A signaling pathway regulates transcriptional activity of SAF-1 by unmasking its DNA-binding domains. *J Biol Chem*. 2003;278(25):22586-22595.
36. Ray A, Yu GY, Ray BK. Cytokine-responsive induction of SAF-1 activity is mediated by a mitogen-activated protein kinase signaling pathway. *Mol Cell Biol*. 2002;22(4):1027-1035.

37. Song J, Ugai H, Nakata-Tsutsui H, et al. Transcriptional regulation by zinc-finger proteins Sp1 and MAZ involves interactions with the same cis-elements. *Int J Mol Med*. 2003;11(5):547-553.
38. Her S, Claycomb R, Tai TC, Wong DL. Regulation of the rat phenylethanolamine N-methyltransferase gene by transcription factors Sp1 and MAZ. *Mol Pharmacol*. 2003;64(5):1180-1188.
39. Pope SH, Fibach E, Sun J, Chin K, Rodgers GP. Two-phase liquid culture system models normal human adult erythropoiesis at the molecular level. *Eur J Haematol*. 2000;64(5):292-303.
40. Ray A, Shakya A, Kumar D, Ray BK. Overexpression of serum amyloid A-activating factor 1 inhibits cell proliferation by the induction of cyclin-dependent protein kinase inhibitor p21WAF-1/Cip-1/Sdi-1 expression. *J Immunol*. 2004;172(8):5006-5015.
41. Andrews NC, Faller DV. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res*. 1991;19(9):2499.
42. Leberbauer C, Boulme F, Unfried G, Huber J, Beug H, Mullner EW. Different steroids co-regulate long-term expansion versus terminal differentiation in primary human erythroid progenitors. *Blood*. 2005;105(1):85-94.
43. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc*. 2007;2(8):1896-1906.
44. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-1372.
45. Lower KM, Hughes JR, De Gobbi M, et al. Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc Natl Acad Sci U S A*. 2009;106(51):21771-21776.
46. Vernimmen D, Lynch MD, De Gobbi M, et al. Polycomb eviction as a new distant enhancer function. *Genes Dev*. 2011;25(15):1583-1588.
47. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360.
48. Diaz A, Park K, Lim DA, Song JS. Normalization, bias correction, and peak calling for ChIP-seq. *Stat Appl Genet Mol Biol*. 2012;11(3):Article 9.
49. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012;22(9):1813-1831.
50. Ramirez F, Ryan DP, Gruning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-165.
51. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
52. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-589.
53. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*. 2014;47:11 12 11-34.
54. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res*. 2015;43(W1):W39-49.
55. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017-1018.

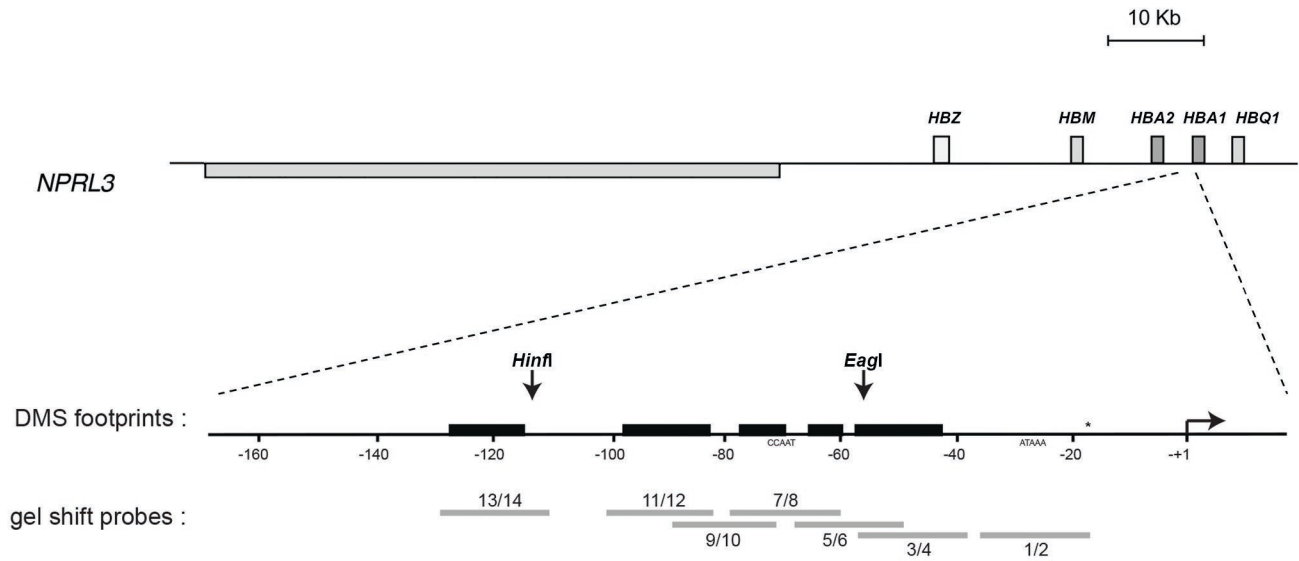
56. Raudvere U, Kolberg L, Kuzmin I, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 2019.
57. Bossone SA, Asselin C, Patel AJ, Marcu KB. MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci U S A.* 1992;89(16):7452-7456.

**Table 1. EMSA Results**

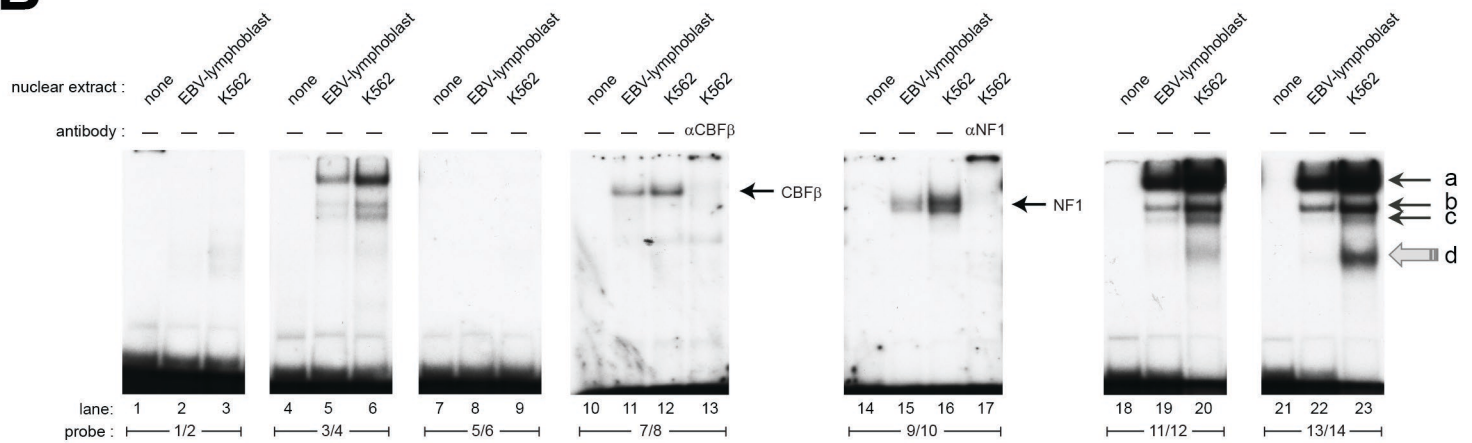
<b>Probe</b>	<b>Location (bp)</b>	<b>Predicted motifs (5)</b>	<b>Gel shift results</b>
1/2	-37/-19	TATA box	No shifted species
3/4	-56/-39	$\alpha$ -IRP -51/-42 site	Shifted species detected in EBV and K562
5/6	-67/-50	Sp1 -61/-56 site	No shift
7/8	-78/-61	CBF -72/-65 site	Shifted species detected in EBV and K562 Binding impaired by $\alpha$ -CBF antibody
9/10	-88/-72	NF1 -85/-72 site	Shifted species detected in EBV and K562 Binding impaired by $\alpha$ -NF1 antibody
11/12	-100/-83		Three shifted species detected in EBV and K562 One shifted species only in K562
13/14	-128/-111	Sp1 -121/-116 site	Three shifted species detected in EBV and K562 One shifted species only in K562

Figure 1

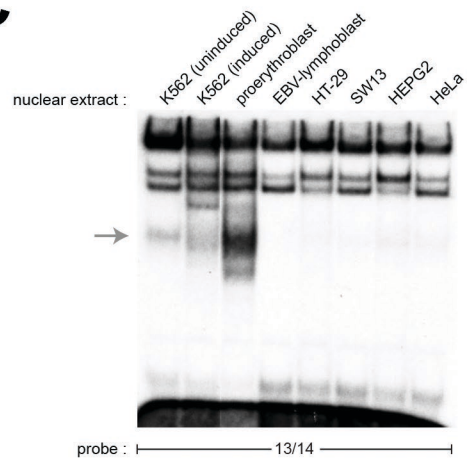
**A**



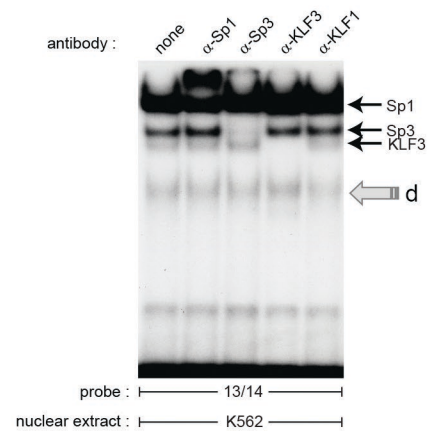
**B**



**C**

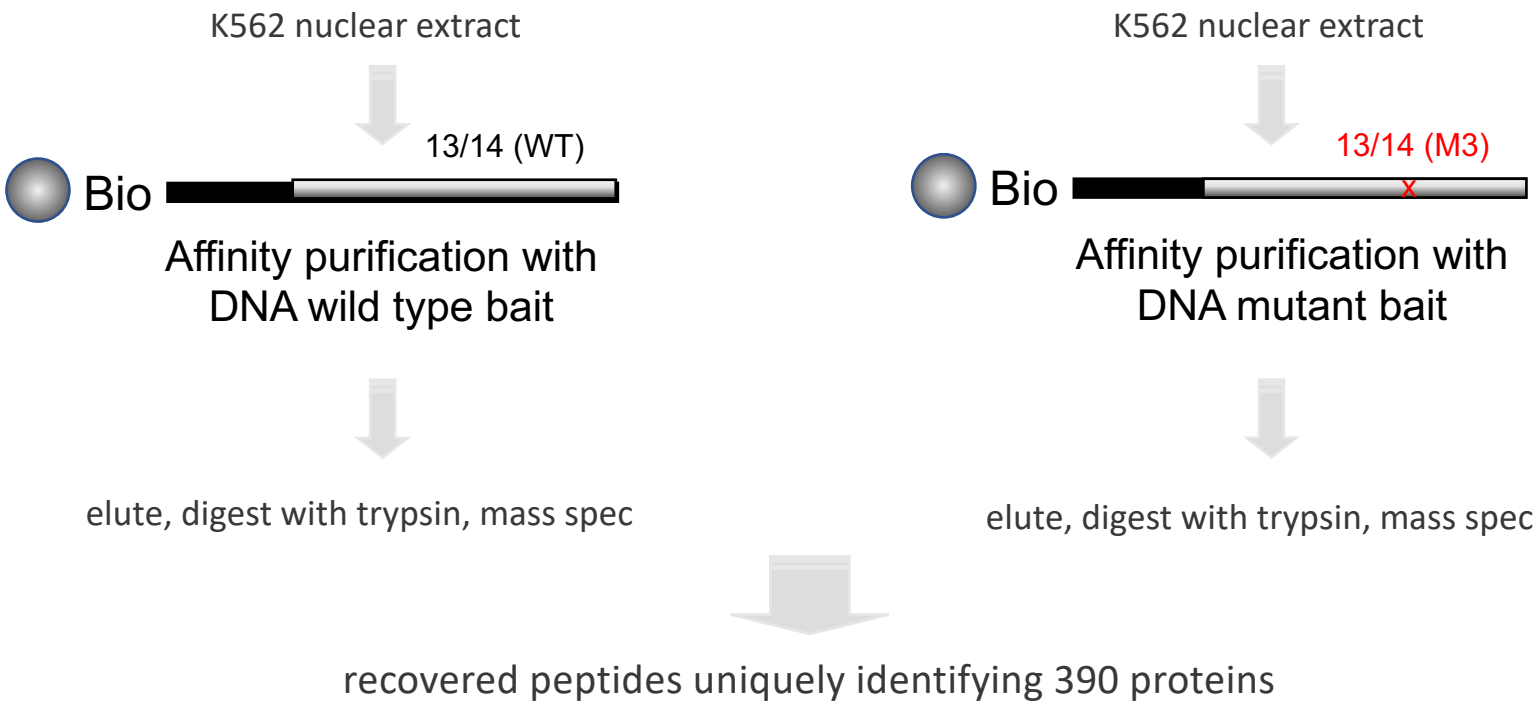


**D**





A

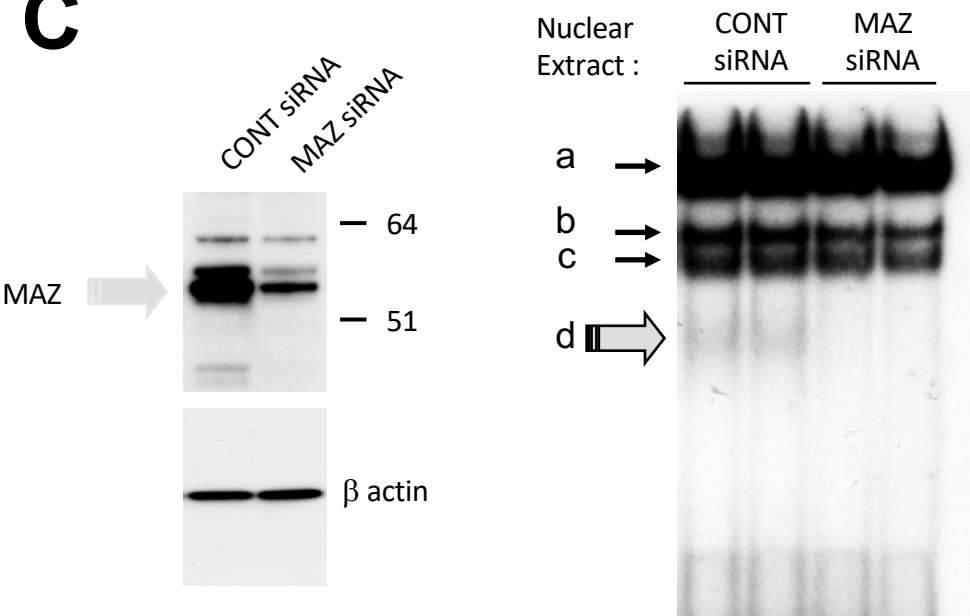


B

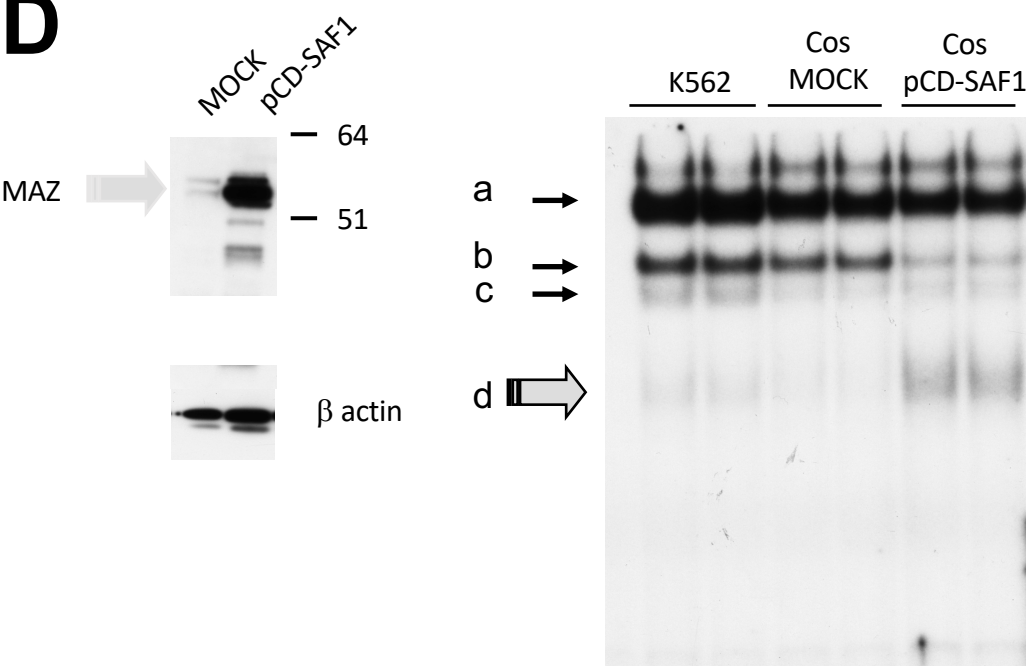
Gene Names	Protein Descriptions	Mol. Weight [kDa]	Enrichment ratio (WT/M3)
FAM139A; FLJ40722; tcag7.730; FLJ00264	Uncharacterized protein ENSP00000291129; Family with sequence similarity 139, member A; Putative uncharacterized protein FLJ40722;	100.9	28.66067999
GNL3; E2IG3; NS	Guanine nucleotide-binding protein-like 3; Nucleolar GTP-binding protein 3; Nucleostemin; E2-induced gene 3 protein; Novel nucleolar protein 47; NNP47	62.0	27.56177032
ANKK1; PKK2; SGK288	Ankyrin repeat and protein kinase domain-containing protein 1; Protein kinase PKK2; X-kinase; Sugen kinase 288; SgK288	84.632	21.63577185
ZCCHC17; PS1D; HSPC243; HSPC251; LDC4	Nucleolar protein of 40 kDa; pNO40; Zinc finger CCHC domain-containing protein 17; Putative S1 RNA-binding domain protein; PS1D protein; Pnn-interacting nucleolar protein	27.6	21.54945043
NOL1	Putative RNA methyltransferase NOL1; Proliferating-cell nucleolar antigen p120; Proliferation-associated nucleolar protein p120; Uncharacterized protein NOL1	92.9	20.70552404
RPS8; OK/SW-cl.83; RP11-269F19.3-003	40S ribosomal protein S8; Ribosomal protein S8	25.2	14.60330352
MAZ	MAZ protein; Myc-associated zinc finger protein; MAZI; Purine-binding transcription factor; Pur-1; ZF87; ZIF87	51.1	14.26242555
SURF6; SURF-6	Surfeit locus protein 6	41.5	14.03328268
HIST1H1D; H1F3	Histone H1.3; Histone H1c	22.4	10.88943545
HIST1H1F; HIST1H1G; HIST1H1H; HIST1H1I; HIST1H1J; HIST1H1K; HIST1H1L; HIST1H1M; HIST1H1N; HIST1H1O; HIST1H1P; HIST1H1Q; HIST1H1R; HIST1H1S; HIST1H1T; HIST1H1U; HIST1H1V; HIST1H1W; HIST1H1X; HIST1H1Y; HIST1H1Z	Histone H1.4; Histone H1b; Histone H1e; Histone H1f; Histone H1g; Histone H1h; Histone H1i; Histone H1j; Histone H1k; Histone H1l; Histone H1m; Histone H1n; Histone H1o; Histone H1p; Histone H1q; Histone H1r; Histone H1s; Histone H1t; Histone H1u; Histone H1v; Histone H1w; Histone H1x; Histone H1y; Histone H1z	21.9	9.228533738

bioRxiv preprint doi: <https://doi.org/10.1101/2020.05.10.087291>; this version posted May 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

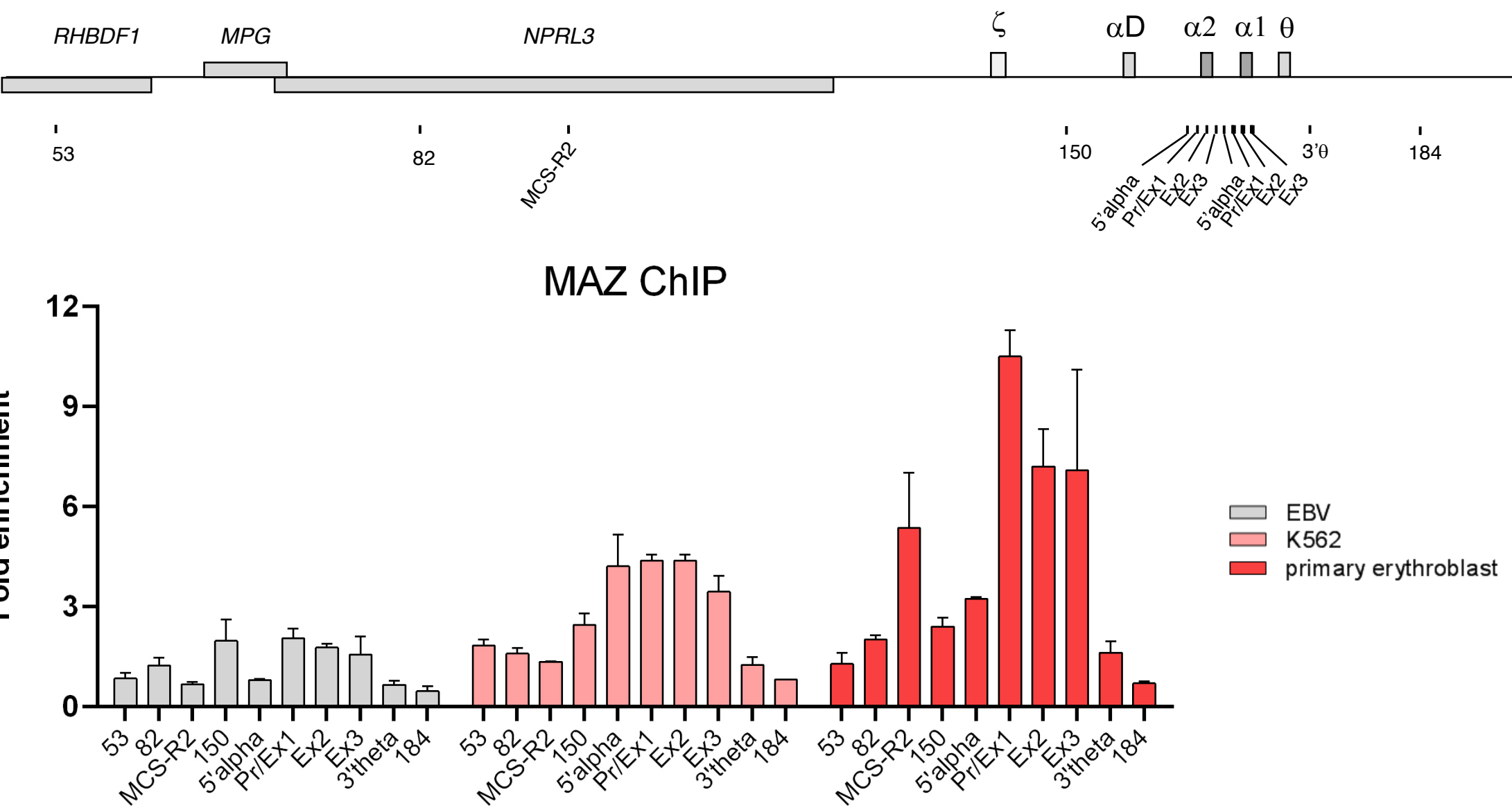
C

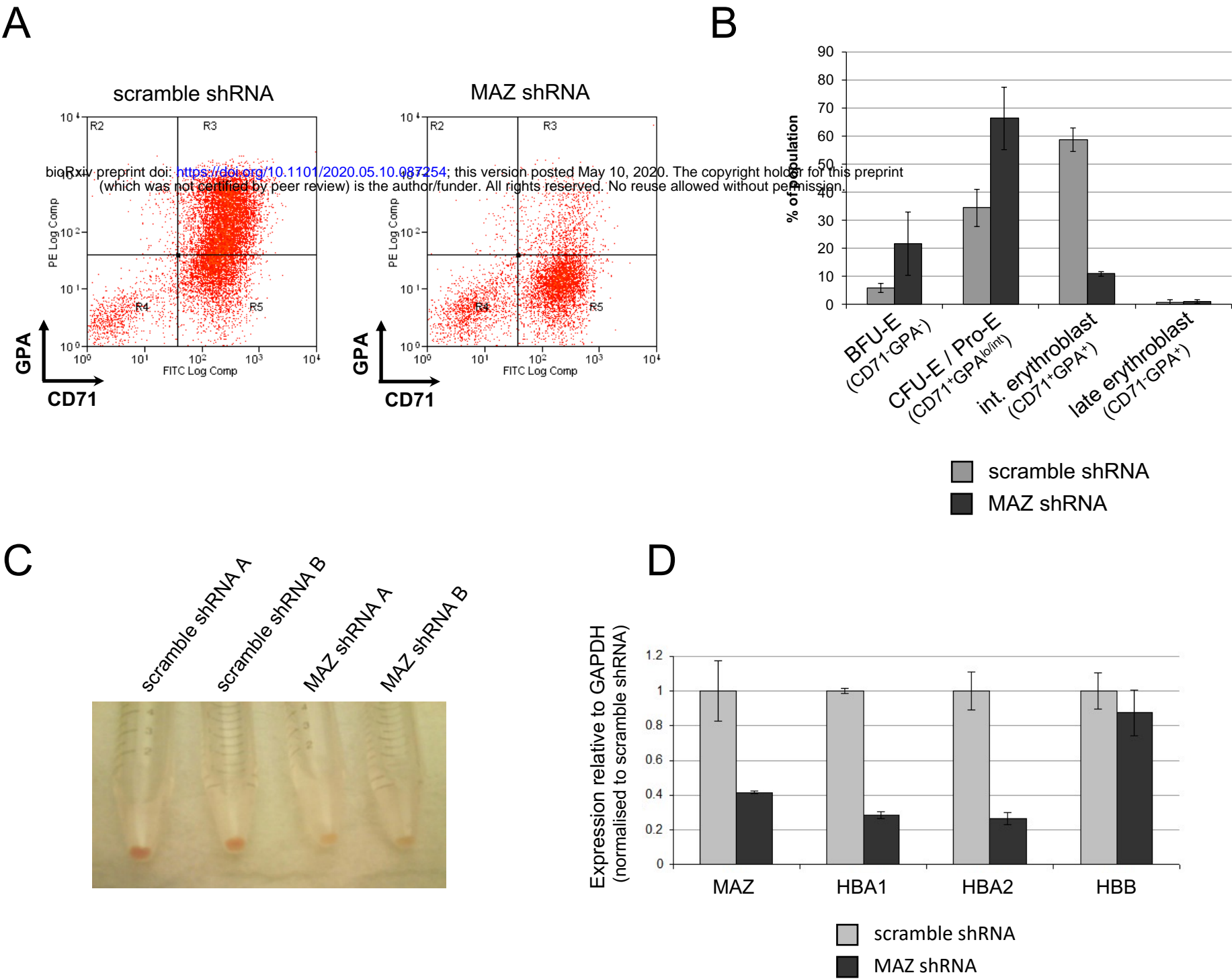


D



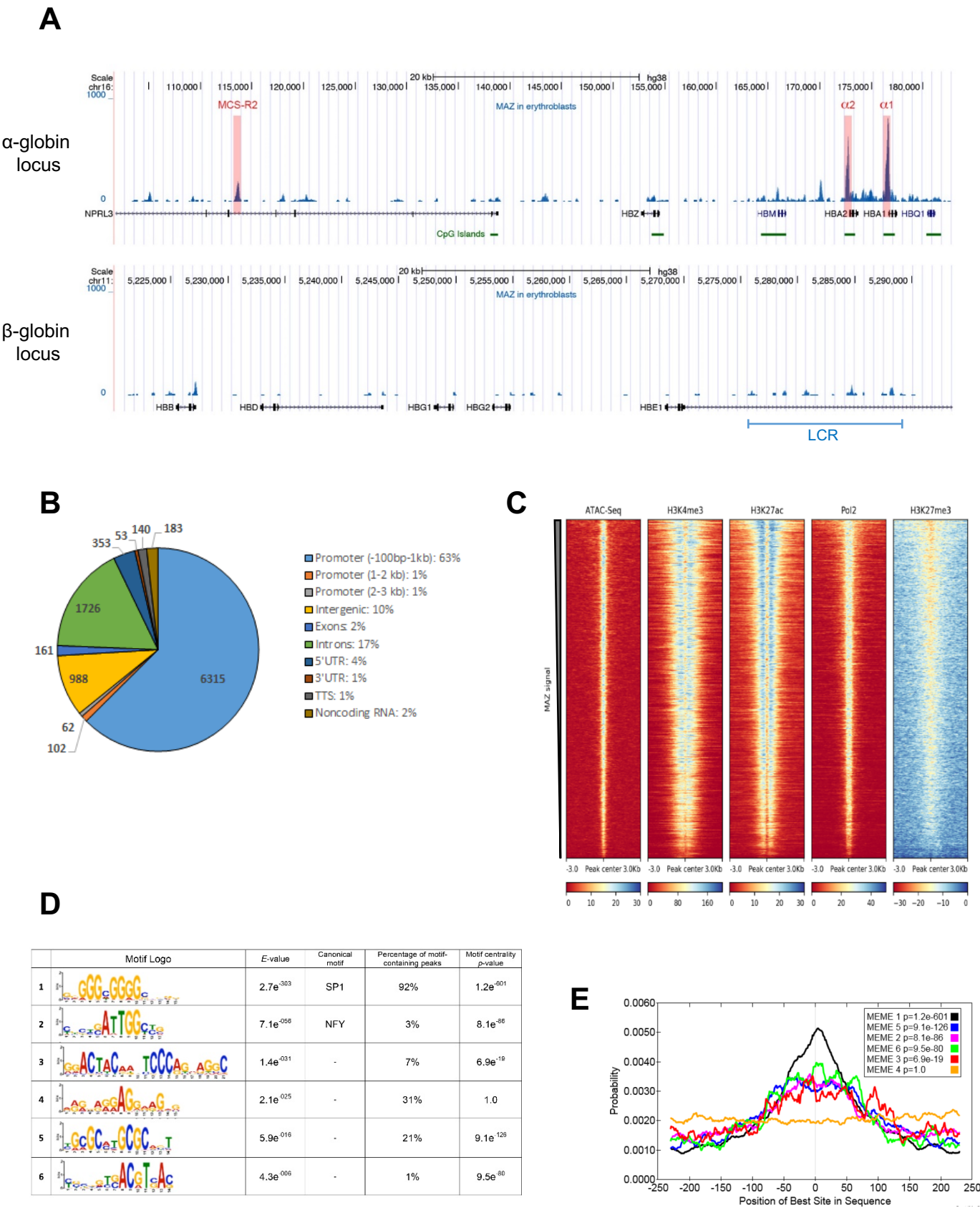
E



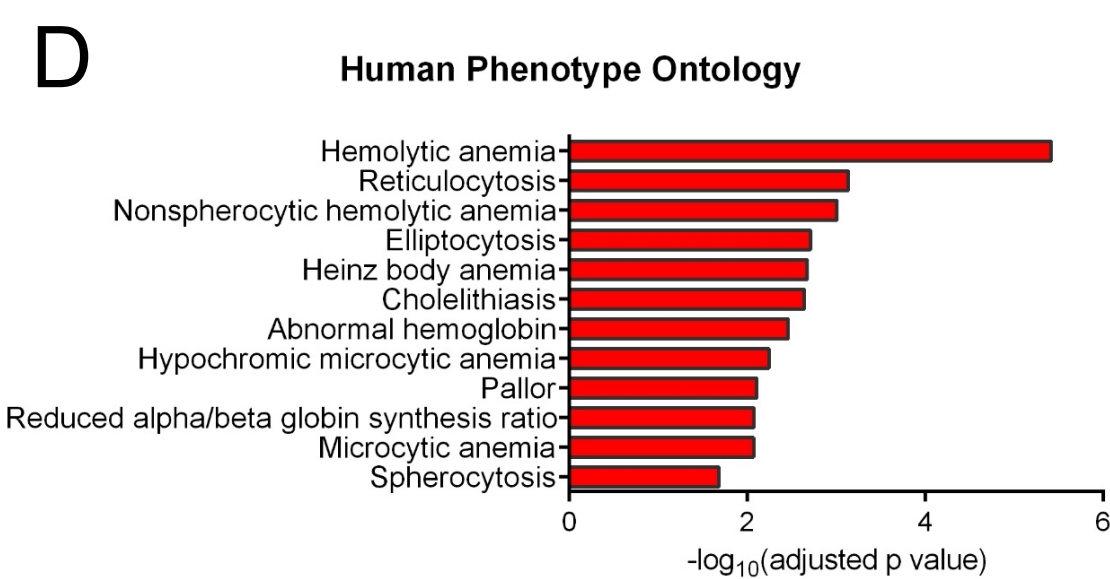
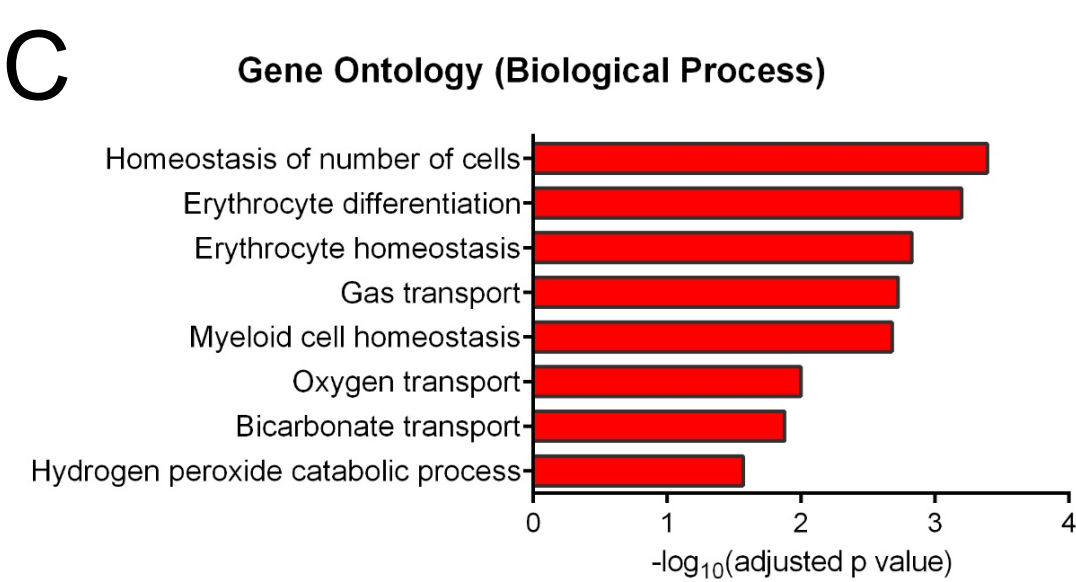
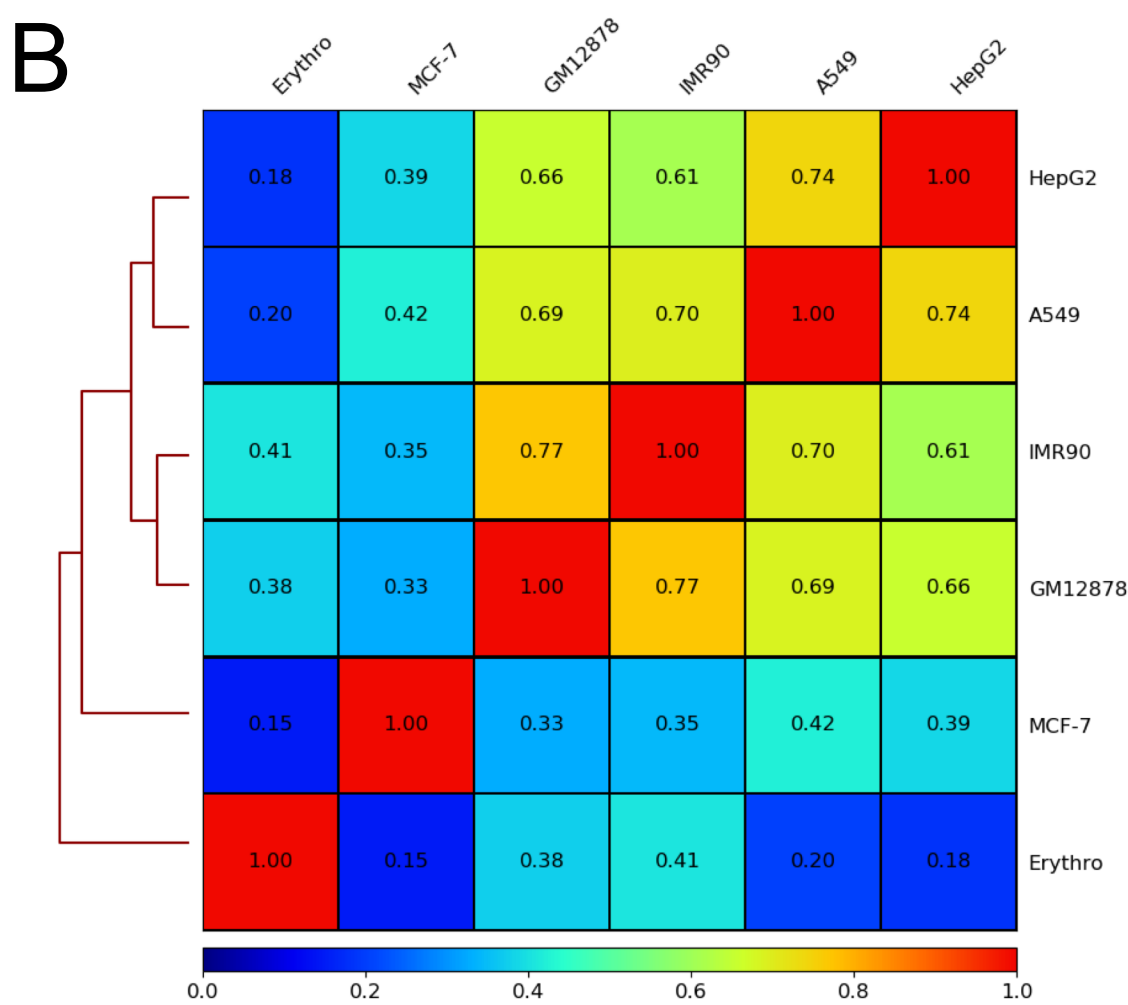
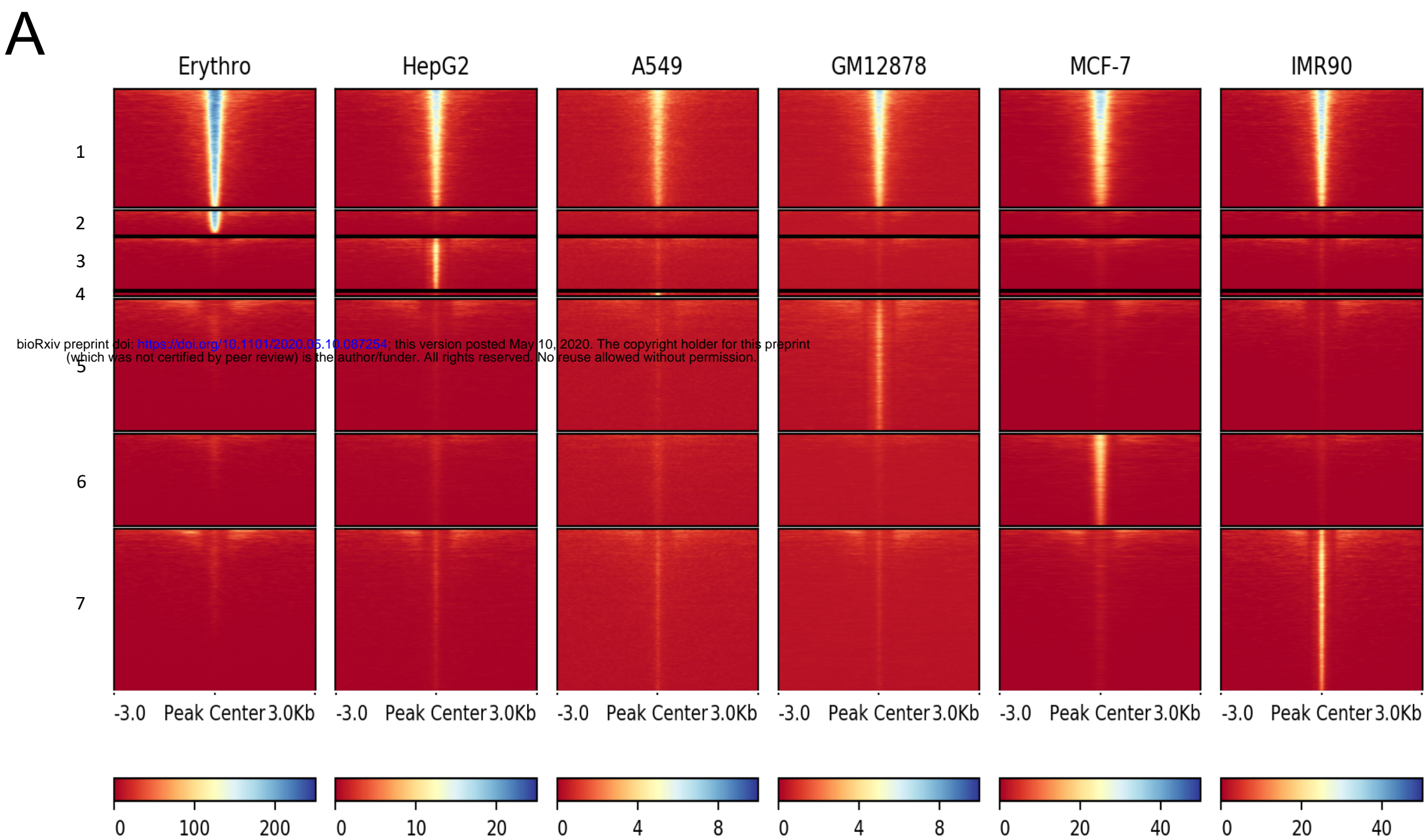


E

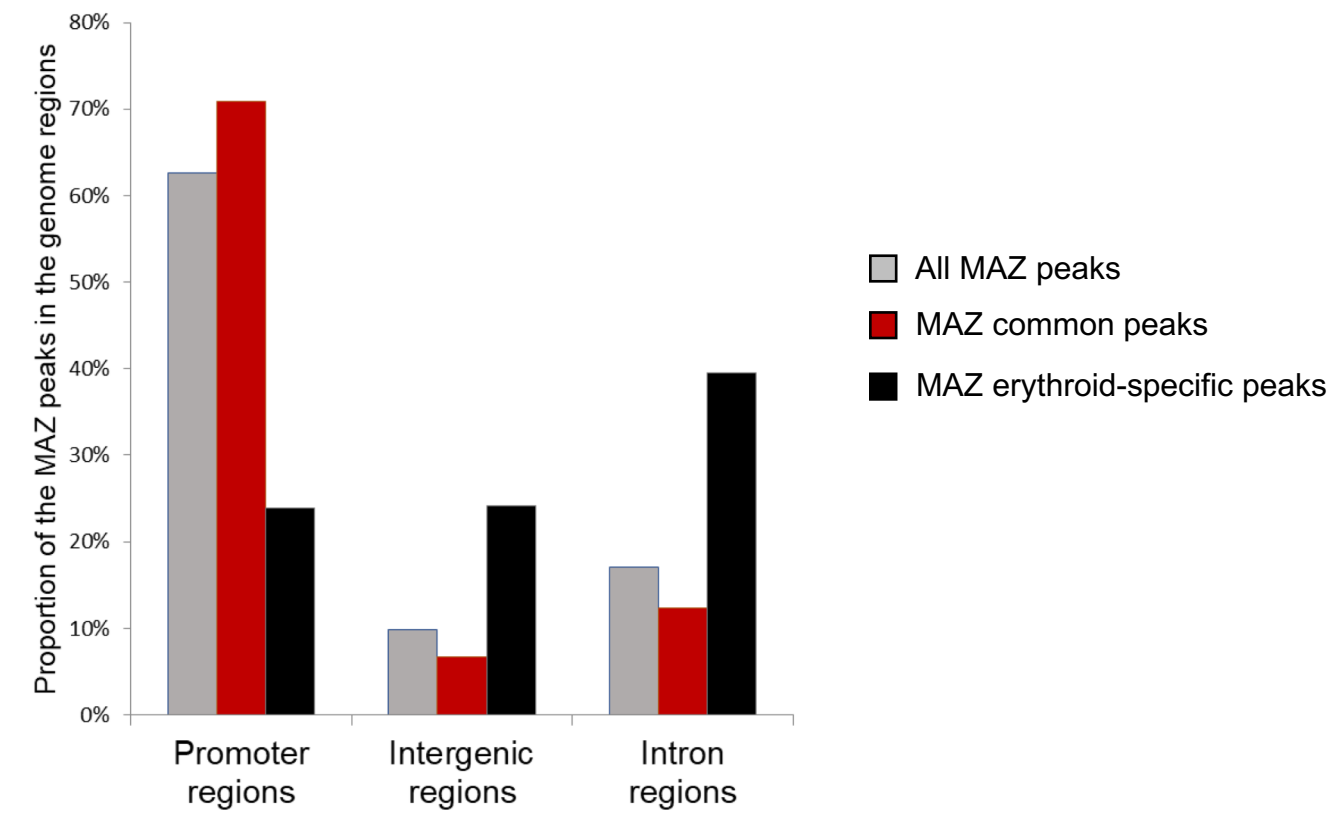
VARIANT	LOCATION IN THE MAZ REGION	ERYTHROID TRAIT	P-VALUE
RS572982482	MAZ promoter	Haematocrit percentage	3.54E-03
		Haemoglobin concentration	7.30E-03
		Mean reticulocyte volume	9.66E-03
		Red blood cell (erythrocyte) count	1.96E-04
		Reticulocyte count	4.85E-04
		Reticulocyte percentage	4.63E-03
RS72798129	intron variant	Mean reticulocyte volume	1.97E-05
		Mean sphered cell volume	8.94E-03
RS11559000	3' UTR variant	D75 Other diseases of blood and blood-forming organs	5.42E-03



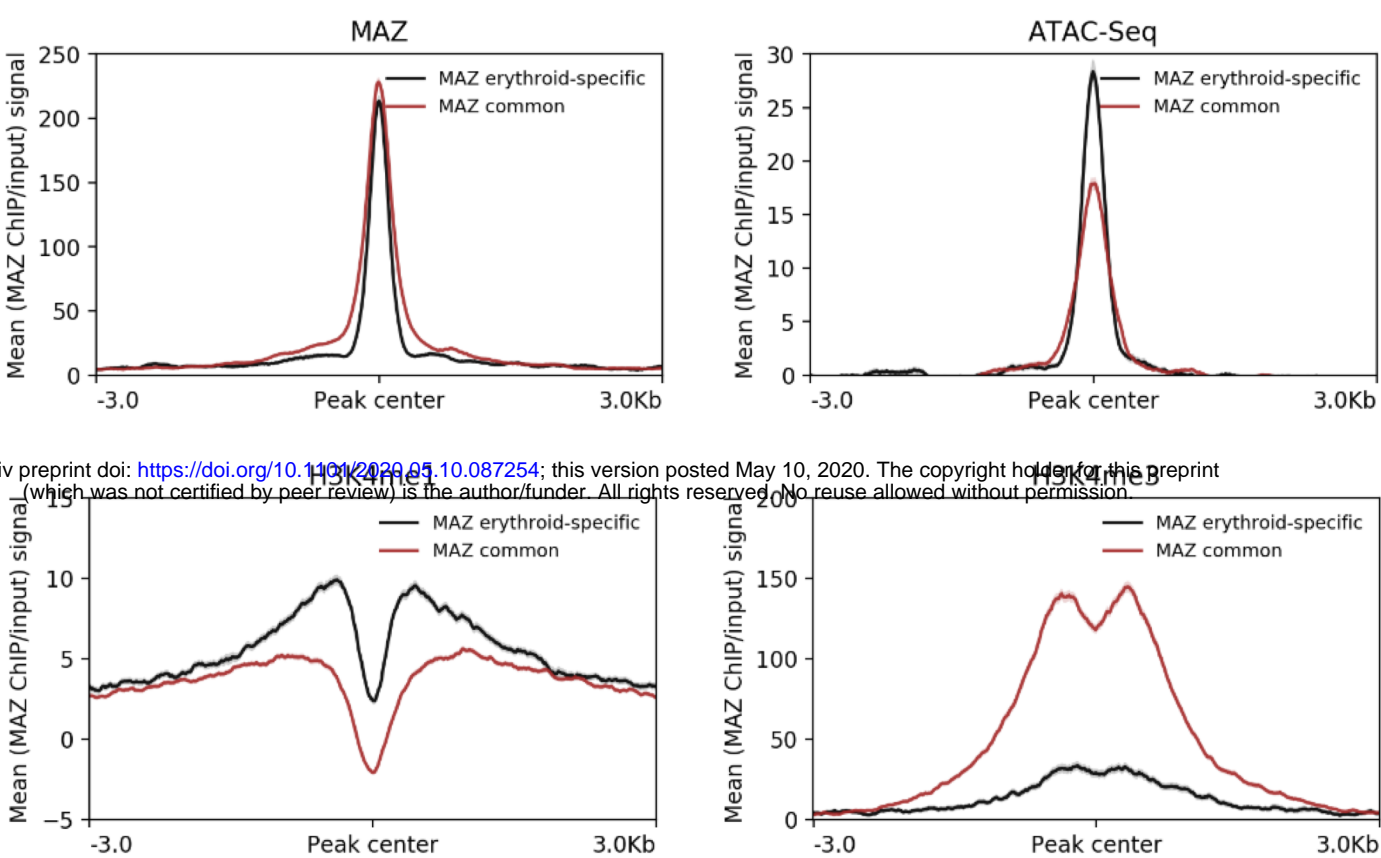




A



B



C

